

Mapping of CO₂ at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2

Dorit M. Hammerling,¹ Anna M. Michalak,^{1,2} and S. Randolph Kawa³

Received 18 October 2011; revised 16 January 2012; accepted 29 January 2012; published 24 March 2012.

[1] Satellite observations of CO₂ offer new opportunities to improve our understanding of the global carbon cycle. Using such observations to infer global maps of atmospheric CO₂ and their associated uncertainties can provide key information about the distribution and dynamic behavior of CO₂, through comparison to atmospheric CO₂ distributions predicted from biospheric, oceanic, or fossil fuel flux emissions estimates coupled with atmospheric transport models. Ideally, these maps should be at temporal resolutions that are short enough to represent and capture the synoptic dynamics of atmospheric CO₂. This study presents a geostatistical method that accomplishes this goal. The method can extract information about the spatial covariance structure of the CO₂ field from the available CO₂ retrievals, yields full coverage (Level 3) maps at high spatial resolutions, and provides estimates of the uncertainties associated with these maps. The method does not require information about CO₂ fluxes or atmospheric transport, such that the Level 3 maps are informed entirely by available retrievals. The approach is assessed by investigating its performance using synthetic OCO-2 data generated from the PCTM/GEOS-4/CASA-GFED model, for time periods ranging from 1 to 16 days and a target spatial resolution of 1° latitude × 1.25° longitude. Results show that global CO₂ fields from OCO-2 observations can be predicted well at surprisingly high temporal resolutions. Even one-day Level 3 maps reproduce the large-scale features of the atmospheric CO₂ distribution, and yield realistic uncertainty bounds. Temporal resolutions of two to four days result in the best performance for a wide range of investigated scenarios, providing maps at an order of magnitude higher temporal resolution relative to the monthly or seasonal Level 3 maps typically reported in the literature.

Citation: Hammerling, D. M., A. M. Michalak, and S. R. Kawa (2012), Mapping of CO₂ at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2, *J. Geophys. Res.*, 117, D06306, doi:10.1029/2011JD017015.

1. Introduction

[2] Atmospheric carbon dioxide (CO₂) is the most important anthropogenic greenhouse gas [Forster *et al.*, 2007]. While data from the existing CO₂ monitoring network have been crucial to gaining important insights into the functioning of the carbon cycle, the mechanisms controlling the interannual variability and the spatial distribution of carbon uptake and emissions are still not fully understood [e.g., Feng *et al.*, 2009; Heimann, 2009; Nevison *et al.*, 2008; Yang *et al.*, 2007]. The accurate prediction and mitigation of climate change requires a better understanding of these processes and the carbon cycle in general [Friedlingstein *et al.*, 2006].

[3] Satellite observations of CO₂, because of their global coverage and high measurement density, offer new

opportunities to improve this understanding. Observations from several satellites are already being used to infer atmospheric CO₂ concentrations, including the Japanese Greenhouse Gases Observing Satellite (GOSAT) [Hamazaki *et al.*, 2004], which is the first satellite dedicated to the measurement of greenhouse gases. NASA's Orbiting Carbon Observatory 2 (OCO-2) is the first American mission designed specifically for making high precision measurement of CO₂ [Crisp *et al.*, 2004], and is expected to be launched in 2015.

[4] Despite their high measurement density, however, satellite CO₂ observations have gaps due to their orbit configurations and due to geophysical limitations such as cloud cover and are subject to substantially higher measurement uncertainties relative to in situ observations. Using statistical techniques to leverage the spatial correlation in the CO₂ concentration field and to predict full-coverage global CO₂ concentration distributions from satellite observations (i.e., creating Level 3 data products) is one way to gain new information about the carbon cycle.

[5] Once derived, such maps can be used for comparison studies with carbon flux estimates coupled with an atmospheric transport model to generate modeled CO₂ fields, or with

¹Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA.

²Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA.

³NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

other available atmospheric measurements. If the satellite-derived Level 3 products were to include rigorous uncertainty measures, such comparisons could be conducted probabilistically, making it possible to assess whether and where, for example, a given set of flux estimates coupled with a specific transport model differ significantly from the satellite-derived Level 3 maps. Such Level 3 products are not intended to be used in inversion studies directly, but instead provide a useful complement to such studies. Beyond point-wise comparisons with individual observations, comparisons with global CO₂ concentration distributions make it possible to identify spatially continuous areas of mismatch, providing indicators for potential discrepancies with other data sets and their dependence on the atmospheric or surface characteristics. Ideally, such comparison studies should be done at high temporal resolution, so that mismatches are not missed through temporal averaging and so that the underlying causes for any mismatches can be tracked in detail. Such comparison studies could, among other applications, inform the growing need to verify and track reported CO₂ emissions [National Research Council, 2010; Nisbet and Weiss, 2010].

[6] There are currently several approaches for creating Level 3 products from CO₂ satellite observations, ranging from simple methods such as spatial and temporal averaging [e.g., Crevoisier et al., 2009; Kulawik et al., 2010; Tiwari et al., 2006] to sophisticated data assimilation approaches [Engelen et al., 2009]. Spatial and temporal averaging entails binning and averaging the data to relatively coarse spatial and temporal grids to obtain smoother maps and to average out the measurement errors. Temporal averaging over months or seasons is commonly applied to satellite data representing properties that vary on seasonal or interannual timescales, such as land cover and phenology. The impact of such temporal averaging on atmospheric CO₂ concentrations, which vary on synoptic timescales, has not been explored. It is obvious, however, that any of the dynamic information, operating at time scales shorter than the temporal averaging time step, is lost. The same applies to spatial variability at scales smaller than the resolution of the spatial averaging grid. Another disadvantage of binning and averaging the data is that the uncertainties associated with the binned data are typically not quantified, which eliminates the option of making probabilistic comparisons.

[7] Data assimilation approaches, on the other hand, require boundary conditions such as carbon flux estimates and transport models to obtain full-coverage global atmospheric CO₂ concentrations. While incorporating this additional information can be powerful, it also implies that the assimilated atmospheric CO₂ fields are sensitive to any misspecification in these prior assumptions. This strong dependence on prior assumptions can especially affect comparison studies: it can be difficult to establish the degree to which apparent similarities or differences between the data-assimilation-derived CO₂ distributions and, for example, coupled biospheric- and atmospheric-transport-model derived CO₂ concentrations are based on similar or dissimilar prior assumptions.

[8] This paper presents and evaluates an alternative method for generating global Level 3 CO₂ products from satellite observations. The method leverages the fact that atmospheric CO₂ concentrations exhibit spatial correlation, by characterizing this spatial correlation and using this information to statistically derive global CO₂ concentrations

and their associated uncertainties. This proposed geostatistical approach accounts for measurement errors and does not require estimates of fluxes or an atmospheric transport model, which is advantageous for comparison studies because the Level 3 products can serve as independent validation data sets.

[9] We use OCO-2 as a prototypical example application for evaluating the method, because making the best use of future OCO-2 observations will represent an important challenge. While the observations will have high precision and a small field of view, their spatial coverage for a given day will be limited. As a result, the length of the time period over which observations are aggregated represents a trade-off between the spatial coverage that the observations can provide and the loss of any information about temporal variability that is masked by combining observations over longer periods. Finding a balance between these effects, and being aware of the consequences of the choice of the length of the aggregation time period, is critical to creating and interpreting global CO₂ maps based on the anticipated data from OCO-2. The presented sample application therefore quantifies the quality of global CO₂ Level 3 products based on simulated OCO-2 observations for time periods ranging from 1 to 16 days.

2. Mapping Methodology

[10] The geostatistical mapping method applied in this study accounts for and exploits the spatial correlation of CO₂ between different locations [e.g., Cressie, 1993; Gelfand, 2010; Chiles and Delfiner, 1999]. First, it infers the spatial covariance structure of the CO₂ concentrations. Second, CO₂ concentrations and associated uncertainties are predicted globally, using the available observations and the spatial covariance structure inferred in the first step. Note that in this study, prediction specifically refers to spatial interpolation of available data, not to temporal prediction.

[11] Alkhaled et al. [2008] showed that global CO₂ concentrations exhibit spatial nonstationarity, such that the expected degree of spatial variability in the CO₂ field itself varies across the globe. For example, CO₂ concentrations over oceans are generally correlated over longer distances than over land. Exploratory analysis of the modeled CO₂ concentrations used here further supports this conclusion, and, as a result, the approach presented here uses a nonstationary statistical framework. The framework chosen is similar to moving window kriging [Haas, 1990], which is, among spatial statistical methods to treat nonstationarity, a rather simple and straightforward approach. From a theoretical point of view, a drawback of moving window kriging is that it does not enforce a globally valid spatial model [e.g., Zhu and Wu, 2010; Chen et al., 2006], but is based on covariance functions that are only valid locally. From a computational point of view, the chosen framework is efficient, as both the estimation of the covariance structure and the prediction of the CO₂ concentrations and their associated uncertainties is executed locally and can be implemented using parallel computing approaches.

2.1. Estimation of Nonstationary Covariance Structure

[12] The global nonstationary covariance structure is estimated using a local semivariogram analysis based on the assumption of local stationarity. The method is similar to

the approach taken by *Alkhaled et al.* [2008]. The spatial covariance structure specific to each location is estimated by using observations in a local neighborhood surrounding this location. The local neighborhood is defined here as a region within 2000-km of each location, following *Alkhaled et al.* [2008], who found such areas to be large enough to capture most of the variability, while being small enough to preserve local phenomena. Further analysis of the neighborhood size conducted in this study confirmed these findings.

[13] Variogram analysis is a tool for quantifying spatial variability as a function of the separation distance between observations. As a first step the raw variogram is calculated:

$$\gamma(h) = \frac{1}{2} [y(x_i) - y(x_j)]^2, \quad (1)$$

where h is the separation distance between locations x_i and x_j , defined as the great circle distance

$$h(x_i, x_j) = r \cos^{-1} \left(\sin \varphi_i \sin \varphi_j + \cos \varphi_i \cos \varphi_j \cos(\lambda_i - \lambda_j) \right), \quad (2)$$

where r is the radius of the Earth and φ_i and λ_i are the latitude and longitude of location x_i , and $y(x_i)$ is the CO₂ value at location x_i . The local variogram analysis is implemented by including all the pairs of observations, where both observations fall within 2000-km of a given location, and a subset of the pairs for which one observation is within 2000-km and the other is further away. The number of pairs in the subset was chosen such that the number of outside observations was a quarter of the number of inside observations. This number is based on a sensitivity analysis for the effect of the selection of outside observations to ensure that the variogram parameters are robustly estimated and do not vary as a function of the randomly selected subset of outside points.

[14] In the second step, a parametric function, the theoretical variogram, is fitted to the raw variogram using non-linear least squares. The function fitted was the exponential variogram function combined with a nugget-effect variogram model given by:

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ \sigma^2 \left(1 - \exp\left(-\frac{h}{l}\right) \right) + \sigma_{nug}^2 & \text{for } h > 0 \end{cases}, \quad (3)$$

where σ^2 and l are the variance and correlation length parameters of the exponential variogram, and σ_{nug}^2 is the nugget variance, which is representative of the retrieval/measurement errors. The choice of the exponential variogram was based on earlier analysis by *Alkhaled et al.* [2008]. The nugget-effect component accounts for the random noise added to the observations in this synthetic-data study to represent the measurement noise (see section 3.2). This variance component is fixed to the variance of the noise added to the observations, and represents the variance of retrieval errors for real data applications. Variogram parameters were estimated for each location on a $1^\circ \times 1.25^\circ$ global grid to match the resolution of the model data used in the analysis (section 3.1), but any convenient resolution could be used with real data from OCO-2 as long as the resolution was sufficiently fine to capture the variability of the data.

[15] The exponential variogram parameters can be used to define an exponential covariance function:

$$C(h) = \sigma^2 \exp\left(-\frac{h}{l}\right), \quad (4)$$

where the parameters σ^2 and l are as defined previously, such that the estimated parameters of the variogram specify the covariance function.

2.2. Local Kriging

[16] Kriging is a minimum variance linear unbiased prediction method for spatial data. Linear refers to the fact that the predicted value at a given location is expressed as a linear combination of the values observed at sampled locations. A notable feature of kriging, differentiating it from simpler interpolation methods such as inverse distance weighting, is that an observation is not only weighted as a function of its distance to the prediction location, but also as a function of its location relative to those of other observations. As such, clustered observations that provide redundant information receive comparatively less weight. Another attractive feature of kriging is that it can account for measurement error. Finally, kriging quantifies the uncertainty in the predicted value.

[17] The linear system that is solved to obtain the weights λ for a single prediction location given observations at n locations is

$$\begin{bmatrix} \mathbf{Q} + \mathbf{R} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix}, \quad (5)$$

where \mathbf{Q} is an $n \times n$ covariance matrix among the n observation locations, as defined in equation (4), \mathbf{R} is an $n \times n$ measurement error covariance matrix among the n observation locations, λ is a $n \times 1$ vector of weights, ν is a Lagrange multiplier and \mathbf{q} is the $n \times 1$ vector of the spatial covariances between an individual prediction location and the observation locations, also defined using equation (4). If the measurement errors are assumed independent between observation locations, as is the case in this work, then \mathbf{R} is a diagonal matrix with the measurement error variance σ_{nug}^2 on the diagonal. The predicted value, \hat{y} , and the prediction uncertainty, σ_y^2 , at the location are:

$$\hat{y} = \lambda^T \mathbf{y} \quad (6)$$

$$\sigma_y^2 = \sigma^2 - \lambda^T \mathbf{q} - \nu, \quad (7)$$

where \mathbf{y} are the observations at the n locations and σ^2 is the variance as shown in equation (4).

[18] ‘Local’ refers to the fact that the covariance parameters used to calculate the spatial covariances are specific to each prediction location, and that only observations within a given neighborhood of the prediction location are considered [e.g., *Haas*, 1990; *Kitanidis*, 1997]. As described in section 2.1, the covariance parameters are derived at each prediction location. Only observations within 2000-km were used in the kriging step, motivated by the shielding effect [*Wackernagel*, 2003]. The validity of the assumption that observations at more than 2000 km have a negligible

influence on the predicted value was verified by comparing the predicted values for increasingly larger neighborhoods (results not shown).

[19] In a small number of cases (less than 0.1% of the prediction locations on average), no observations were available within 2000 km, and the kriging procedure could not be applied. In these cases a simple imputation technique was applied using the predicted value and uncertainty of the closest location where the local kriging procedure could be executed.

3. Study Design and Data

[20] The study was designed to evaluate how well global CO₂ concentrations can be reconstructed from satellite observations using a geostatistical mapping method. The specific emphasis was on recreating global CO₂ concentrations based on future OCO-2 observations for short time periods ranging from 1 day to one repeat cycle (i.e., 16 days).

[21] OCO-2 is scheduled for launch in 2015, and is a replacement for OCO, which failed upon launch, and was to be NASA's first satellite mission dedicated to observing atmospheric CO₂. Some of the most noteworthy features of OCO-2 are the sensitivity to the near-surface CO₂ abundance, the measurement footprint of about 3 km², and an anticipated measurement precision of 1 ppm once soundings are averaged over regional scales [Crisp *et al.*, 2004]. OCO-2 will be part of NASA EOS Afternoon constellation (A-train) [L'Ecuyer and Jiang, 2010], which flies in a sun-synchronous polar orbit with a 16-day repeat-cycle.

3.1. Simulated OCO-2 CO₂ Observations

[22] The atmospheric CO₂ field is simulated using the PCTM/GEOS-4/CASA-GFED (referred to simply as PCTM in the discussion that follows) atmospheric model coupled with biospheric, biomass burning, oceanic, and anthropogenic CO₂ flux estimates [Kawa *et al.*, 2004, 2010]. This model uses analyzed meteorological fields to drive both the biospheric flux and atmospheric transport. The model grid is 1° × 1.25° × 28 vertical levels with hourly output. The PCTM/GEOS-4 model has been widely tested, and has shown good results in carbon cycling comparison studies [e.g., Kawa *et al.*, 2004; Law *et al.*, 2008; Parazoo *et al.*, 2008]. CO₂ mixing ratios in the lowest 20 vertical layers of the model (up to 40 mbar) were pressure-averaged to simulate the vertical sensitivity of OCO-2. Prospective OCO-2 sounding locations were determined by overlaying the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) [Winker *et al.*, 2003] track on the CO₂ field for each day within a given repeat-cycle. The CALIPSO track was used because this satellite is also part of NASA's A-train constellation, and CALIPSO flies only minutes apart from the orbit planned for OCO-2. Differences in the OCO-2 slant path and glint location offset were not accounted for, so the CALIPSO track is only a close approximation to the true OCO-2 track.

[23] The presence of clouds and aerosols will impede the retrieval of atmospheric CO₂ concentrations, leading to gaps in the OCO-2 observations along the satellite track. To represent the presence of these gaps in a realistic manner, the combined cloud and aerosol optical depth (532 nm) from

the version 2.01 5-km CALIPSO data was used to identify locations on the track where the total cloud and aerosol optical depth was below 0.3, which is a conservative estimate of the maximum optical depth that will allow for the successful retrieval of CO₂ [D. M. O'Brien, personal communication, 2011]. Using this approach to account for clouds and aerosols has the advantage of matching the CALIPSO data with the PCTM output in time, which allows for a more realistic representation of the cloud-aerosol-CO₂ distribution relative to using probabilistic cloud and aerosol masks based on seasonal averages. The CALIPSO along-track horizontal resolution of 5 km was matched with the coarser PCTM 1° × 1.25° horizontal resolution by considering a model grid box visible if at least one CALIPSO measurement with a combined optical depth of less than 0.3 fell within the grid box. Figures 1d–1f show typical patterns and amounts of visible locations (at the PCTM grid resolution) for 1-day, 4-day and 16-day time periods.

[24] OCO-2's footprint of approximately 3 km² will be much finer than the PCTM horizontal resolution used in this study, and the simulated OCO-2 observations used here therefore most closely resemble a setup where the true observations would be pre-averaged to the PCTM/GEOS resolution of 1° × 1.25°. This setup has implications for the measurement error characteristics. Having multiple OCO-2 soundings within a PCTM grid box reduces the measurement error associated with the average CO₂ value in the grid box relative to the uncertainty of a single sounding. The relative reduction of the measurement error is a function of the number, spatial configuration, and measurement error correlation of the soundings within a grid box. The procedure and the assumptions made to account for multiple soundings within a grid box and characterize the measurement error associated with a grid box are described in detail in section 3.2.

[25] Some other features associated with the finer resolution of the true OCO-2 observations are not directly assessed in this study; namely, the computational aspects associated with the number of OCO-2 observations, and the possibility of capturing CO₂ variability on very fine scales. OCO-2's fine resolution leads to a large total number of observations, up to hundreds of thousands each day, which could cause computational problems for traditional geostatistical gap-filling methods. The local covariance estimation and kriging setup described here (section 2), however, are well suited to handling a large amount of observations, and have been specifically designed to do so.

3.2. Experimental Setup

[26] The primary goals of the experiment were to evaluate (1) the overall performance of the proposed approach, and (2) how the temporal resolution, which is the length of the time period over which observations are accumulated to make a single map, affected the quality of the resulting map. The quality of the obtained maps was evaluated by comparing these maps and their inferred uncertainties to the full model data, which were time-averaged over the period of the observations used to create the maps. The details of the comparison measures are discussed in sections 4.2 and 4.3. The experiment was specifically designed to assemble observations in a manner that is realistic for satellite observations: the simulated observations were not taken from a

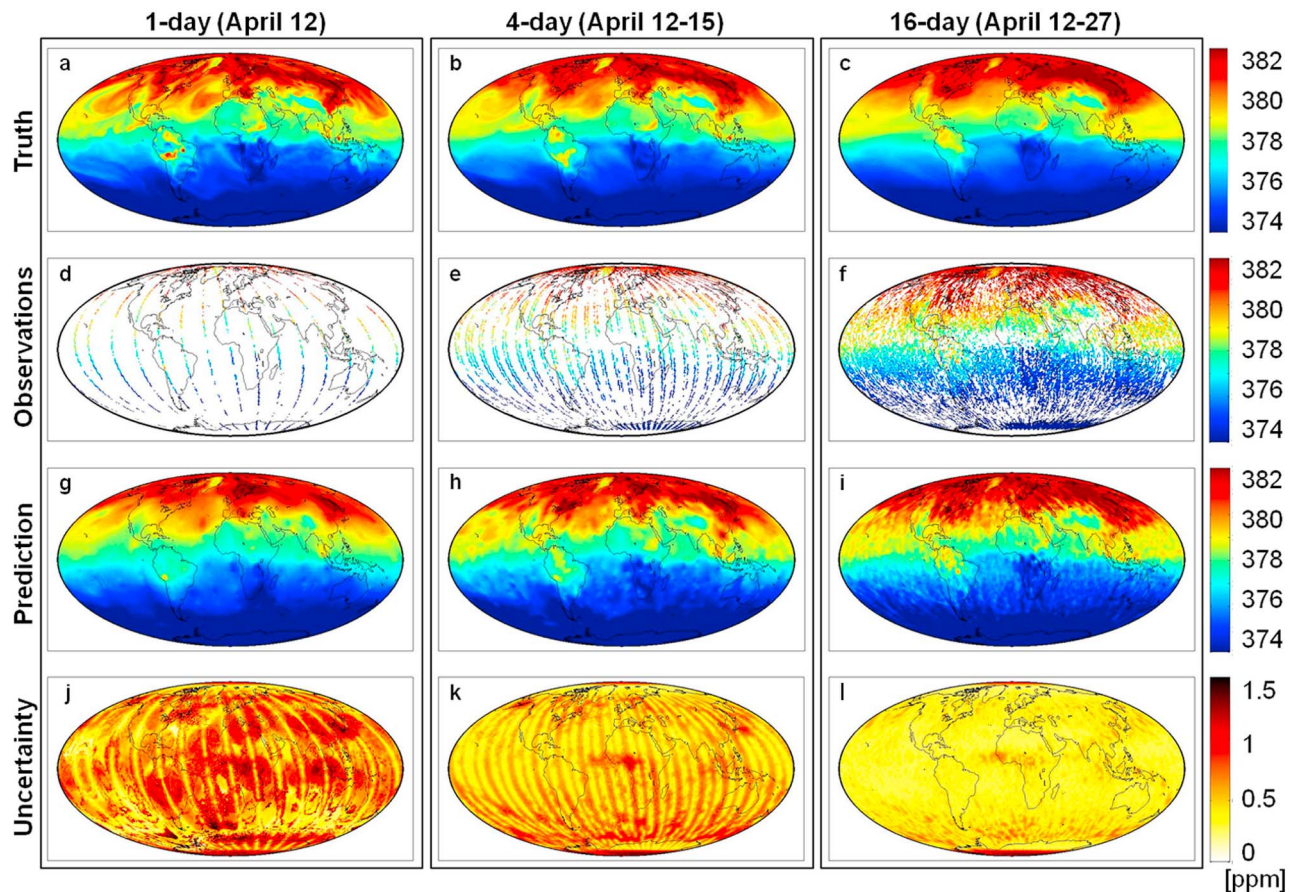


Figure 1. PCTM/GEOS-4 CO₂ model output (“Truth”), simulated OCO-2 observations (“Observations”), gap-filled predictions (“Prediction”) and estimated prediction uncertainties (“Uncertainty”) expressed as a standard deviation for a 1-day, a 4-day and a 16-day time period in April 2006. The observation locations and values are derived by overlaying CALIPSO track and cloud and aerosol information on PCTM/GEOS-4 CO₂ from individual days in the repeat cycle. The gap-filled predictions are shown for the medium measurement noise scenario with the covariance estimated from observations.

time-averaged CO₂ concentration field, but were sampled from individual days at the hour nearest the local overpass time (approximately 1330 h). For example, the observations shown in Figure 1f are sampled from 16 different days of PCTM output, corresponding to the actual day for each sounding. This way of simulating observations results in a field that represents an aggregation of observations from different days rather than temporally averaged observations. The true field, however, is the full 3-D model output time-averaged over the aggregation period. Figures 1a–1c provide an example of the 1-day, 4-day and 16-day true fields.

[27] In addition to the temporal resolution, the season, measurement noise level, and data used in the covariance estimation were also varied to evaluate the approach.

[28] The temporal resolutions evaluated were 16-day, 8-day, 4-day, 2-day and 1-day intervals. These lengths were chosen to 1) identify the shortest time period for which meaningful global CO₂ maps can be obtained from OCO-2 observations and to 2) quantify the effect of temporal resolution, and thereby the amount of data and temporal variability within the time period, on prediction performance.

[29] To explore the impact of seasonality on the heterogeneity of the atmospheric CO₂, cloud and aerosol distributions,

one month was used as representative of each season (January, April, July, September). For example, April and July featured higher variability in the CO₂ concentration field than January and September.

[30] A range of assumed measurement error levels was selected based on the expected performance of OCO-2. The levels of measurement noise are based on a single sounding expected OCO-2 measurement error standard deviation of 1.5 ppm [Crisp *et al.*, 2004; D. M. O’Brien, personal communication], but accounting for the fact that a single model grid box may contain multiple OCO-2 soundings. Because nearby OCO-2 observations will likely have correlated errors, the effective measurement error at the grid scale will be higher relative to a case with independent measurement errors. The effective measurement error standard deviation at the grid scale is a function of the number of soundings and their spatial configuration within the grid cell, as well as the spatial scale over which the measurement errors are correlated. The number and spatial configuration of OCO-2 soundings was estimated by examining the range of the number, and the spatial configuration, of CALIPSO measurements with optical depths of less than 0.3 falling within a PCTM grid box. Simulated OCO-2 observations and

measurement error correlation ranges from a few kilometers to a few hundred kilometers were used in a side study to determine the effective measurement noise at the scale of the PCTM model. Based on these results (not shown), a range of grid-scale measurement error standard deviations was applied, which were 0.2 ppm for the low level, 0.5 ppm for the medium level and 1 ppm for the high level. For all noise levels, the measurement errors were assumed to be independent zero-mean and normally distributed when applied at the scale of the model ($1^\circ \times 1.25^\circ$), and a random sample of such errors was added to the observations drawn from PCTM/GEOS-4. The nugget variance σ_{nug}^2 , and therefore the diagonal elements of the matrix \mathbf{R} , defined in section 2.1 was thereby equal to the variances of these measurement errors (i.e., $(0.2 \text{ ppm})^2$, $(0.5 \text{ ppm})^2$ and $(1 \text{ ppm})^2$).

[31] The fourth factor in the experimental setup was the data used in the covariance estimation (see section 2.1). Two cases were investigated. In the first case, the time-averaged full model data (e.g., Figures 1a–1c), i.e., the “truth,” were used to derive the covariance parameters. In the second case, only the available observations (e.g., Figures 1d–1f) were used. Using the time-averaged full-model data represents an idealized, but not possible scenario, where the covariance structure could be derived from the full time-averaged CO₂ concentrations. Clearly, having the true concentrations available to estimate the covariance structure for gap-filling CO₂ is not feasible and would defy the need to gap-fill, but this choice has been made to serve as an upper bound for any possible improvement over the observation-based covariance structure. The idea is that any alternative to using the observations themselves to quantify the covariance structure would be at best as good as having the “truth.”

4. Results and Discussion

4.1. Qualitative Features of the Spatial Predictions

[32] The characteristics of the Level 3 maps as a function of the length of the examined time period, amount of measurement noise, and data used in the covariance estimation were similar across seasons (Figure 2). As expected, seasons with smoother CO₂ fields yielded better Level 2 maps. The large-scale features of the global CO₂ fields could be reproduced for all examined scenarios. It is surprising and encouraging that, even for the 1-day periods, the information content of the observations is sufficient to recover the main characteristics of the CO₂ field (Figure 1).

[33] Figures 1g–1i provide an example of gap-filled estimates for 1-day, 4-day and 16-day periods in April, which was the season with the most heterogeneous CO₂ field. As expected, some small-scale features are lost, especially in the 1-day maps, if they are not captured by observations. For example, the area of high CO₂ values over the northern part of South America is not well portrayed in the 1-day Level 3 map (Figure 1g). However, as is discussed in detail in section 4.3, the prediction uncertainties for the shorter time periods adequately reflect the true uncertainty of these predictions. So, while the predictions cannot recreate the small-scale features in areas missing observations, the associated prediction uncertainties are higher in these areas, and therefore reflect this lack of information.

[34] The smoothness of the predicted fields varies as a function of the length of the examined time period. The 1-day

and 2-day predictions are generally smoother than the truth, whereas for the longer time periods, most notably the 16-day periods, the predictions are less smooth than the truth. This can be seen in the undulating structure of the 16-day prediction map for April (Figure 1i). There are two different causes for this change of smoothness with temporal resolution. The reason for which the 1-day and 2-day maps appear smoother than the true fields is a general consequence of interpolating sparse data. The reason for which prediction maps for longer time periods appear less smooth than the truth is a consequence of unaccounted-for temporal variability in the CO₂ field, as reflected in the available observations. For the longer time periods, this effect dominates because the data density is relatively high, as is the amount of temporal variability that is captured by these observations. This temporal variability is introduced into the gap-filled maps because they combine observations from different days (see section 3.2). The amount of temporal variability that is captured by the observations increases with the time span over which observations are combined, and its effect therefore becomes more pronounced for longer time periods. In a spatial-only (compared to a spatiotemporal) geostatistical setup such as the one used here, the temporal variability translates to a perceived spatial variability on small-scales (Figures 1h and 1i). In contrast, the corresponding true fields shown in Figures 1b and 1c represent the temporal average over the examined time period and are smoother. The strength of this effect is further affected by the degree of measurement noise: the undulating structure is less pronounced in high measurement noise scenarios, where the measurement noise masks the temporal variability (see section 4.2).

4.2. Prediction Accuracy

[35] Figure 2 presents the Root Mean Square Prediction Error (RMSPE) for all modeled scenarios. RMSPE is a measure of the difference between the true and predicted CO₂ values. The overall range of RMSPE was 0.20 to 0.63 ppm CO₂; the lowest value resulted from a 16-day period in September and the highest from a 1-day period in July. Longer time periods and seasons with lower CO₂ variability generally had better prediction accuracies, i.e., lower RMSPE. These overall trends, however, depend on the level of measurement noise. For the 1-day periods, prediction accuracies improve as the measurement noise decreases. This ordered relationship is less evident in the 2-day periods, where the low and medium noise scenarios have similar prediction accuracies. For the 4-day periods, the medium measurement noise cases have the best prediction accuracies, but the differences are less pronounced than for other temporal resolutions. For the 8-day periods, the relationship between measurement noise and prediction accuracies starts to reverse: lower measurement noise is associated with worse prediction accuracies. This effect becomes fully evident in the 16-day periods, where higher measurement noise scenarios consistently feature the best prediction accuracies for all seasons, because temporal variability dominates the error. Overall, higher measurement noise decreases the prediction accuracies for shorter time periods, but, counter-intuitively, improves them for longer time periods.

[36] This effect is due to the presence of temporal variability in the observations. As discussed in sections 3.2

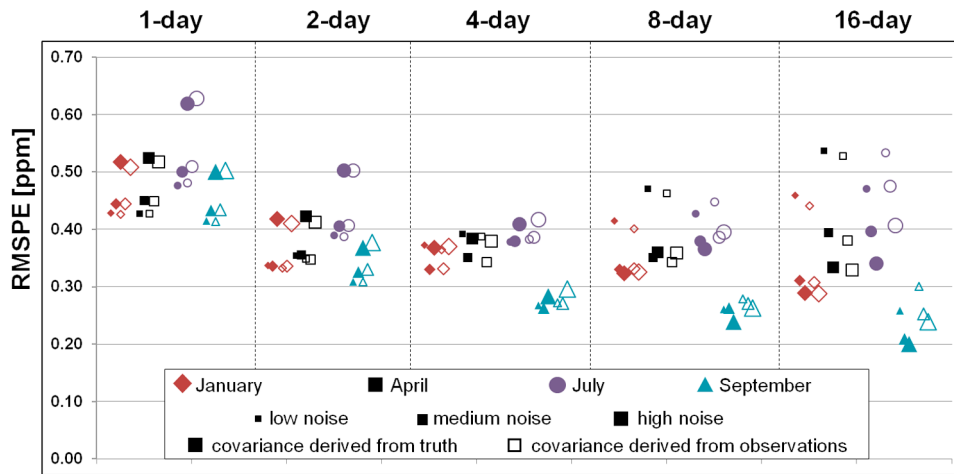


Figure 2. Root mean square prediction error (RMSPE) for 1-day, 2-day, 4-day, 8-day and 16-day maps for January, April, July and September 2006. Symbols of different colors and shapes represent different seasons. Symbol sizes indicate different measurement noise levels (see section 3.2). Filled and empty symbols represent results obtained when the covariance structure is derived from the true field or from the available observations, respectively.

and 4.1, temporal variability in the CO₂ distribution is captured by the observations by combining measurements from multiple days. This variability, however, is not accounted for directly by the spatial mapping approach presented here, which treats observations from different days as if they had been sampled from a static field. As a result, predictions may follow observations too closely to accurately represent the averaged field. This effect is alleviated in cases with high measurement noise, because the geostatistical modeling framework provides leeway for the predictions to deviate from the observations to a degree consistent with the measurement error. In this way, accounting for high measurement noise implicitly also allows the method to cope with observed temporal variability. There is also an interaction between measurement error and the heterogeneity of the CO₂ field. Seasons with more spatial heterogeneity also exhibited more temporal variability. As a result, seasons with smoother CO₂ fields also yielded better prediction accuracies for longer time periods and lower measurement noise, relative to more heterogeneous seasons.

[37] A second measure of prediction accuracy is the percentage of locations where the predicted values deviate from the truth by more than 1 ppm, with results (Figure 3) consistent with the RMSPE results. The lowest percentage was observed for the 16-day period with high measurement error in September, where only 0.2% of the predicted CO₂ values deviate from the truth by more than 1 ppm, and the highest percentage was for a 1-day period with high measurement error in July, where 9% of the predicted values deviated from the truth by more than 1 ppm. As seen previously, lower measurement noise improved the prediction accuracy for short periods, whereas higher measurement noise improved accuracies for long periods. For all cases, however, the percentages are quite low, indicating high accuracy predictions by the proposed method.

[38] Surprisingly and encouragingly, whether the covariance structure was derived from the model data averaged over the examined time period, i.e., the truth that we are

trying to estimate (e.g., Figures 1a–1c), or from the available observations (e.g., Figures 1d–1f), had little impact on the prediction accuracies (see Figure 2). This indicates that good predictions can be obtained without the need for prior information about the covariance structure of the underlying field. This was surprising especially for the shorter time periods, which had more limited observations, and indicated that data over short periods still contain enough information about the spatial variability of the underlying field to yield accurate predictions. The only scenarios where deriving the covariance structure from the full model output (i.e., from prior information other than the observations) improved the prediction accuracies were some of the longer time periods; namely the 8-day and 16-day time periods for July and September. As described in detail in section 3.2, the averaged field over the time period investigated was defined as the truth, while the observations were aggregated from individual days, and thus did not come from an averaged field. Therefore, the observations come from a more variable field than the truth, and that variability is reflected in the estimated covariance structures, which translates into somewhat less accurate prediction accuracies for the longer time periods.

4.3. Prediction Uncertainty

[39] An attractive feature of geostatistical mapping is that each predicted CO₂ value is accompanied by a prediction uncertainty that is quantified without knowledge of the true distribution. The prediction uncertainty for a given location is a function of the location and number of observations surrounding the location, and the degree of spatial variability in the CO₂ field in the vicinity of the estimation location (equations (5) and (7)). In general, more homogeneous areas with dense observations will have lower prediction uncertainty.

[40] Figures 1j–1l provide an example of the prediction uncertainties, as obtained from the approach implemented here, for the 1-day, 4-day and 16-day periods for April.

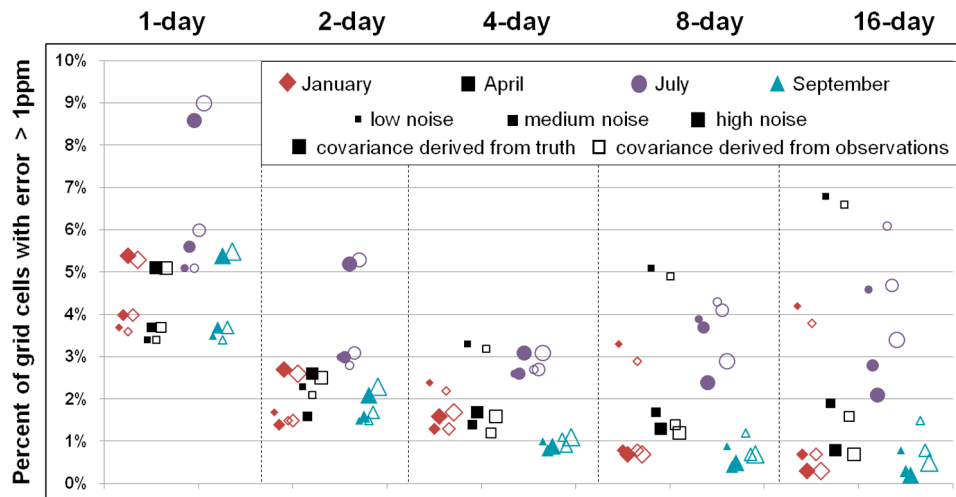


Figure 3. Percentage of locations where the predicted values deviate from the truth by more than 1 ppm for 1-day, 2-day, 4-day, 8-day and 16-day time periods for January, April, July and September 2006. Symbols of different colors and shapes represent different seasons. Symbol sizes indicate different measurement noise levels (see section 3.2). Filled and empty symbols represent results obtained when the covariance structure is derived from the true field or from the available observations, respectively.

The 1-day prediction uncertainties show clear evidence of the dependence of the prediction uncertainties on the location of observations. Locations that lie close to the satellite orbit path feature low uncertainties, while areas further away from the satellite paths have increasingly larger uncertainties. The 16-day period, which has a large number of observations distributed over the globe, features overall lower prediction uncertainties compared to the shorter time periods. Even for a temporal resolution of 16-days, however, some areas with few observations, such as West Africa and the polar regions, have higher prediction uncertainties.

[41] Accurately assessing the uncertainty associated with predictions is valuable regardless of the ultimate use of the maps, but it is especially critical when the global gap-filled CO₂ predictions are compared to data from other sources such as model predictions. Realistic prediction uncertainties allow for probabilistic comparisons in addition to evaluating the best estimates. In order to assess how representative the prediction uncertainties were of the true uncertainty, the percentage of estimation locations where the truth fell outside of the estimated value ± 3 standard deviations (as calculated from the prediction uncertainty) was evaluated. The optimal percentage for this measure depends on assumptions about the underlying statistical distribution of the data. As a guiding value, under the assumption of a normal distribution, this percentage should be approximately 0.3%. While achieving this exact value is not the goal, because the approach does not assume that the underlying distribution is Gaussian, we have assessed whether the percentage outside of ± 3 standard deviations is reasonably low.

[42] Figure 4 shows the percentage of locations falling outside of ± 3 standard deviations of the prediction uncertainty for all investigated scenarios. The most striking feature of this figure is how the percentage dramatically increases with the length of the examined time period for low-noise scenarios, while the percentage stays low for high-noise scenarios. This feature is in accordance with the

finding, discussed in detail in section 4.2, that high measurement noise can mask the temporal variability that is not otherwise accounted for by the spatial mapping. The 1-day scenarios, where temporal variability is minimal, have their lowest percentages for the low measurement noise cases. Starting with the 2-day temporal resolution, however, low measurement noise results in increasingly higher percentages of true values falling outside of ± 3 standard deviations. For the high measurement error cases, accounting for the noise implicitly also accounts for the temporal variability and the percentages falling outside of ± 3 standard deviations remain low.

[43] The effect of the method used for deriving the covariance on the prediction uncertainty depends on the averaging time, but is overall small. For the 1-day periods, using only the observations to derive the covariance structure is clearly suboptimal. This is reflected in the higher percentage of locations falling outside of ± 3 standard deviations (Figure 4). This is not such a clear-cut case, however, for the longer time periods. While the truth-derived covariance structure still has the advantage of being based on a full field without gaps, it is possible that the observation-derived covariance results in improved prediction uncertainties by capturing some of the temporal variability present in the observations. This is indeed the case for some of the 8-day and 16-day. These improvements, however, were rather small compared to the differences caused by the varying degrees of measurement noise.

[44] Overall, the prediction uncertainties are able to describe the true uncertainty accurately. This is an especially encouraging finding for the short time periods, because it indicates that satellite observations can be used to derive global CO₂ distributions with accurate uncertainties for time periods as short as one day. For longer time periods and low measurement noise scenarios, it is important to assess and incorporate the temporal variability resulting from the aggregation of observations to avoid an underestimation of

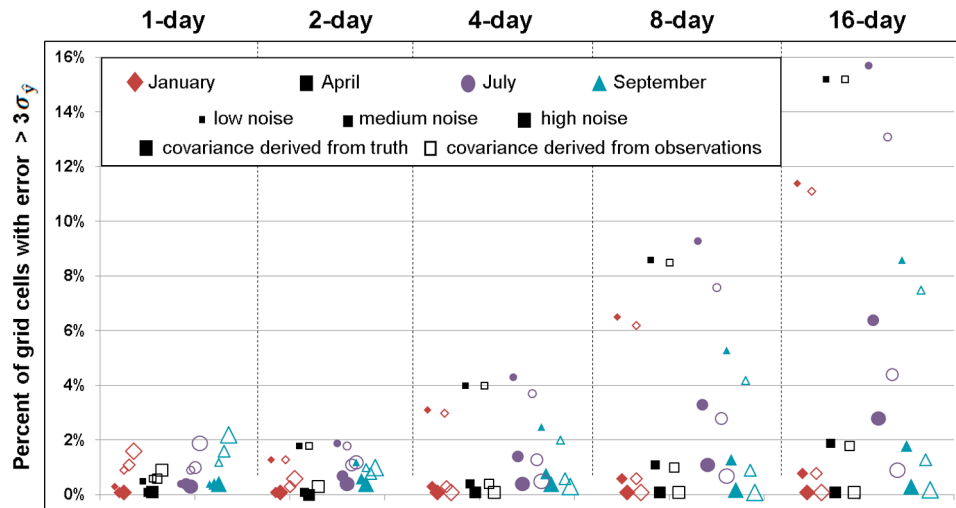


Figure 4. Percentage of locations where the predicted values deviate from the truth by more than ± 3 standard deviations of the prediction uncertainty (equation (7)) for 1-day, 2-day, 4-day, 8-day and 16-day time periods for January, April, July and September 2006. Symbols of different colors and shapes represent different seasons. Symbol sizes indicate different measurement noise levels (see section 3.2). Filled and empty symbols represent results obtained when the covariance structure is derived from the true field or from the available observations, respectively.

the uncertainty. This could be achieved by either calculating the temporal variability and explicitly accounting for it in a geostatistical model, or by developing a spatiotemporal mapping approach.

4.4. Implications for the Generation of Level 3 Maps

[45] It is challenging to construct Level 3 CO₂ maps that represent an average over a given time period using observations obtained on individual days, because CO₂ fields change with time. Ideally, the temporal resolution at which maps are obtained optimizes the mapping performance and provides maps that are representative over the shortest time period possible so as to capture the dynamics of the CO₂ distribution. The choice of temporal resolution thus defines a trade-off between having sufficient observations for adequate spatial coverage, while minimizing the impact of temporal variability. The findings described in sections 4.1 to 4.3 quantify this trade-off and provide guidance for choosing a temporal resolution for creating Level 3 products from satellite CO₂ observations from OCO-2.

[46] When choosing a temporal resolution, results show that a key question is how the measurement noise compares to the temporal variability present in the estimated field. As a general guideline, the larger the measurement noise, the more advantageous it is to combine observations over a longer time period. For observations with low measurement noise, however, choosing a temporal resolution coarser than four days leads to decreased overall prediction performance.

[47] Choosing a high temporal resolution, and thereby sacrificing spatial coverage by observations in favor of minimal temporal variability, can lead to surprisingly benign consequences in prediction performance. Even for the 1-day and 2-day periods, the RMSPE are only on the order of 0.5 ppm and 0.4 ppm, respectively. Furthermore, the accompanying prediction uncertainties accurately reflect the true uncertainty of the predictions.

[48] Overall, for OCO-2 like observations, a temporal resolution of 4-days has the most robust prediction performance for varying seasons and measurement noise levels. Higher measurement noise shifts the optimal prediction performance toward lower temporal resolutions (i.e., longer time periods), while lower measurement noise shifts it toward higher temporal resolutions (i.e., shorter time periods).

[49] Whether the covariance structure is derived from the model data averaged over the examined time period (i.e., the truth that we are trying to estimate) or from the observations had very little impact on the quality of the Level 3 prediction and uncertainty maps. This finding strongly supports the use of observations for deriving the covariance structure, thereby avoiding the need for prior assumptions about the spatial structure of the CO₂ field.

5. Conclusions

[50] High spatiotemporal resolution global Level 3 CO₂ products obtained from satellite observations offer new opportunities for gaining a better understanding of the distribution and dynamic behavior of atmospheric CO₂. Ideally, these Level 3 products should cover time periods that are short enough to preserve the synoptic dynamics of atmospheric CO₂ concentrations. Knowledge of the uncertainties associated with statistically derived Level 3 maps makes it possible to probabilistically evaluate CO₂ flux and atmospheric transport models, which can help identify potential areas for improvement in model formulation and parameterization.

[51] A common method for the generation of Level 3 maps is to obtain an aggregated field by spatial binning and averaging over long periods, which results in a loss of spatial resolution and dynamic information. While making monthly or seasonal maps might be adequate for more static properties (e.g., land cover, phenology), creating CO₂ maps over these long time periods hides the dynamics of the global

CO₂ concentration field, which are critical to improving our understanding of the carbon cycle. Such averaged fields also typically lack quantitative uncertainty measures.

[52] The method presented in this study makes it possible to map CO₂ for time scales more consistent with the synoptic dynamics of CO₂, and provides a measure of the uncertainty associated with predictions. This proposed method makes minimal assumptions, namely that the atmospheric CO₂ concentration exhibit spatial correlation, and that the statistical characteristics of this correlation can be inferred from the observations. Using only the observations themselves to infer the covariance structure eliminates the need to introduce any a priori assumptions about the distribution of atmospheric CO₂ concentrations, which in turn renders the methodology more useful for comparison purposes.

[53] The methodology was used to evaluate Level 3 products derived from OCO-2-like data for time periods ranging from 1 to 16 days, with the dual goal of verifying the proposed method's performance and of identifying the optimal temporal resolution for Level 3 CO₂ products. The results indicate that global CO₂ concentrations can be predicted from OCO-2 satellite observations for time periods much shorter than a full repeat cycle. Even one-day prediction maps reproduce the large-scale features of the atmospheric CO₂ distribution and have realistic uncertainty bounds. Temporal resolutions of 2 to 4 days proved to have the most robust prediction performances over a wide variety of tested scenarios. The aggregation of observations over longer time periods introduces temporal variability that limits prediction performance, especially for scenarios where the measurement noise is low compared to the degree of temporal variability in the underlying CO₂ field.

[54] **Acknowledgments.** This material is based upon work supported by the National Aeronautics and Space Administration under grant NNX08AJ92G issued through the Research Opportunities in Space and Earth Sciences (ROSES) Carbon Cycle Science program. We would like to thank Denis O'Brien, Igor Polonsky, Alanood Alkhaled, Abhishek Chatterjee, Noel Cressie, Matthias Katzfuss, and Amy Braverman for their helpful comments and contributions.

References

- Alkhaled, A., A. M. Michalak, S. R. Kawa, S. Olsen, and J. Wang (2008), A global evaluation of the regional spatial variability of column integrated CO₂ distributions, *J. Geophys. Res.*, **113**, D20303, doi:10.1029/2007JD009693.
- Chen, L., M. Fuentes, and J. Davis (2006), Spatial-temporal statistical modeling and prediction of environmental processes, in *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*, edited by J. S. Clark and A. E. Gelfand, pp. 121–144, Oxford Univ. Press, New York.
- Chiles, J., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, 695 pp., John Wiley, Hoboken, N. J.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, rev. ed., 900 pp., John Wiley, Hoboken, N. J.
- Crevoisier, C., A. Chédin, H. Matsueda, T. Machida, R. Armante, and N. A. Scott (2009), First year of upper tropospheric integrated content of CO₂ from IASI hyperspectral infrared observations, *Atmos. Chem. Phys.*, **9**(14), 4797–4810, doi:10.5194/acp-9-4797-2009.
- Crisp, D., et al. (2004), The Orbiting Carbon Observatory (OCO) mission, *Adv. Space Res.*, **34**(4), 700–709, doi:10.1016/j.asr.2003.08.062.
- Engelen, R. J., S. Serrar, and F. Chevallier (2009), Four-dimensional data assimilation of atmospheric CO₂ using AIRS observations, *J. Geophys. Res.*, **114**, D03303, doi:10.1029/2008JD010739.
- Feng, L., P. I. Palmer, H. Bösch, and S. Dance (2009), Estimating surface CO₂ fluxes from space-borne CO₂ dry air mole fraction observations using an ensemble Kalman filter, *Atmos. Chem. Phys.*, **9**(8), 2619–2633, doi:10.5194/acp-9-2619-2009.
- Forster, P., et al. (2007), Changes in atmospheric constituents and in radiative forcing, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 130–234, Cambridge Univ. Press, Cambridge, U. K.
- Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison, *J. Clim.*, **19**(14), 3337–3353, doi:10.1175/JCLI3800.1.
- Gelfand, A. E. (2010), *Handbook of Spatial Statistics*, 607 pp., CRC Press, Boca Raton, Fla.
- Haas, T. C. (1990), Kriging and automated variogram modeling within a moving window, *Atmos. Environ., Part A*, **24**(7), 1759–1769, doi:10.1016/0960-1686(90)90508-K.
- Hamazaki, T., Y. Kaneko, and A. Kuze (2004), Carbon dioxide monitoring from the GOSAT satellite, paper presented at XXth ISPRS Conference, ISPRS, Istanbul, Turkey.
- Heimann, M. (2009), Searching out the sinks, *Nat. Geosci.*, **2**(1), 3–4, doi:10.1038/ngeo405.
- Kawa, S. R., D. J. Erickson III, S. Pawson, and Z. Zhu (2004), Global CO₂ transport simulations using meteorological data from the NASA data assimilation system, *J. Geophys. Res.*, **109**, D18312, doi:10.1029/2004JD004554.
- Kawa, S. R., J. Mao, J. B. Abshire, G. J. Collatz, X. Sun, and C. J. Weaver (2010), Simulation studies for a space-based CO₂ lidar mission, *Tellus, Ser. B*, **62**(5), 759–769, doi:10.1111/j.1600-0889.2010.00486.x.
- Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications to Hydrogeology*, 272 pp., Cambridge Univ. Press, New York, doi:10.1017/CBO9780511626166.
- Kulawik, S. S., et al. (2010), Characterization of Tropospheric Emission Spectrometer (TES) CO₂ for carbon cycle science, *Atmos. Chem. Phys.*, **10**(12), 5601–5623, doi:10.5194/acp-10-5601-2010.
- Law, R. M., et al. (2008), TransCom model simulations of hourly atmospheric CO₂: Experimental overview and diurnal cycle results for 2002, *Global Biogeochem. Cycles*, **22**, GB3009, doi:10.1029/2007GB003050.
- L'Ecuier, T., and J. Jiang (2010), Touring the atmosphere aboard the A-Train, *Phys. Today*, **63**(7), 36, doi:10.1063/1.3463626.
- National Research Council (2010), *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements*, 124 pp., Natl. Acad. Press, Washington, D. C.
- Nevison, C. D., N. M. Mahowald, S. C. Doney, I. D. Lima, G. R. van der Werf, J. T. Randerson, D. F. Baker, P. Kasibhatla, and G. A. McKinley (2008), Contribution of ocean, fossil fuel, land biosphere, and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO₂, *J. Geophys. Res.*, **113**, G01010, doi:10.1029/2007JG000408.
- Nisbet, E., and R. Weiss (2010), Top-down versus bottom-up, *Science*, **328**(5983), 1241–1243, doi:10.1126/science.1189936.
- Parazoo, N. C., A. S. Denning, S. R. Kawa, K. D. Corbin, R. S. Lokupitiya, and I. T. Baker (2008), Mechanisms for synoptic variations of atmospheric CO₂ in North America, South America and Europe, *Atmos. Chem. Phys.*, **8**(23), 7239–7254, doi:10.5194/acp-8-7239-2008.
- Tiwari, Y. K., M. Gloor, R. J. Engelen, F. Chevallier, C. Rödenbeck, S. Körner, P. Peylin, B. H. Braswell, and M. Heimann (2006), Comparing CO₂ retrieved from Atmospheric Infrared Sounder with model predictions: Implications for constraining surface fluxes and lower-to-upper troposphere transport, *J. Geophys. Res.*, **111**, D17106, doi:10.1029/2005JD006681.
- Wackernagel, H. (2003), *Multivariate Geostatistics: An Introduction With Applications*, 3rd rev. ed., 387 pp., Springer, Berlin.
- Winker, D. M., J. R. Pelon, and M. P. McCormick (2003), The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds, *Proc. SPIE Int. Soc. Opt. Eng.*, **4893**(1), 1–11, doi:10.1117/12.466539.
- Yang, Z., R. A. Washenfelder, G. Keppel-Aleks, N. Y. Krakauer, J. T. Randerson, P. P. Tans, C. Sweeney, and P. O. Wennberg (2007), New constraints on Northern Hemisphere growing season net flux, *Geophys. Res. Lett.*, **34**, L12807, doi:10.1029/2007GL029742.
- Zhu, Z., and Y. Wu (2010), Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data, *J. Comput. Graph. Stat.*, **19**(1), 74–95, doi:10.1198/jcgs.2009.07123.

D. M. Hammerling, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

S. R. Kawa, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

A. M. Michalak, Department of Global Ecology, Carnegie Institution for Science, 260 Panama St., Stanford, CA 94305, USA. (michalak@stanford.edu)